Random walk on sparse random digraphs

Charles Bordenave, Pietro Caputo, Justin Salez

September 29, 2015

Abstract

A finite ergodic Markov chain exhibits *cutoff* if its distance to equilibrium remains close to its initial value over a certain number of iterations and then abruptly drops to near 0 on a much shorter time scale. Originally discovered in the context of card shuffling (Aldous-Diaconis, 1986), this remarkable phenomenon is now rigorously established for many reversible chains. Here we consider the non-reversible case of random walks on sparse directed graphs, for which even the equilibrium measure is far from being understood. We work under the configuration model, allowing both the in-degrees and the out-degrees to be freely specified. We establish the cutoff phenomenon, determine its precise window and prove that the cutoff profile approaches a universal shape. We also provide a detailed description of the equilibrium measure.
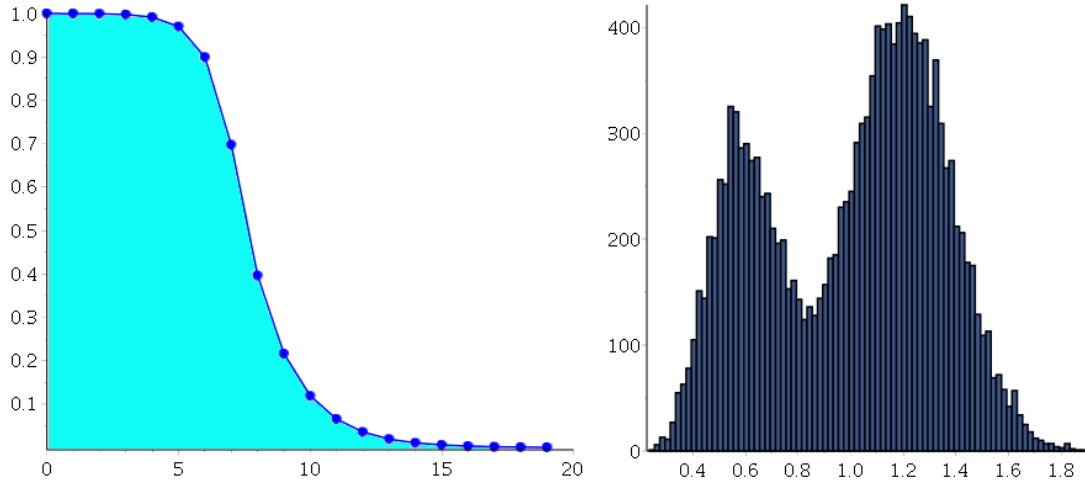
Figure 1: Distance to equilibrium along time (left) and histogram of the vertex weights multiplied by n under the equilibrium measure (right) for the random walk on a random digraph with $n = 5000 + 5000 + 5000$ vertices with respective degrees $(d^+, d^-) = (3, 2), (3, 4)$ and $(4, 4)$.

1 Introduction

Given two sequences of positive integers $(d_i^-)_{1 \leq i \leq n}$ and $(d_i^+)_{1 \leq i \leq n}$ with equal sum m, we construct a directed multigraph G with in-degrees $(d_i^-)_{1 \leq i \leq n}$ and out-degrees $(d_i^+)_{1 \leq i \leq n}$ as follows. We formally equip each vertex $i \in V := \{1, \ldots, n\}$ with a set E_i^+ of d_i^+ *tails* and a set E_i^- of d_i^- *heads*. We then simply choose a tail-to-head bijection $\omega \colon \bigcup_i E_i^+ \to \bigcup_i E_i^-$ (the *environment*), and interpret each coordinate $\omega(e) = f$ as an arc ef from the vertex of e to that of f (loops and multiple edges are allowed). Our interest is in the *Random Walk* on the resulting directed graph (digraph) G, i.e. the discrete-time Markov chain with state space V and transition matrix

$$P(i,j) \quad := \quad \frac{1}{d_i^+} \operatorname{card}\left\{e \in E_i^+ : \omega(e) \in E_j^-\right\}.$$

Starting from $i \in V$, the probability that the walk is at $j \in V$ at time $t \in \mathbb{N}$ can be expanded as

$$P^t(i,j) \quad = \quad \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}), \tag{1}$$

where the sum ranges over all *directed paths* $\mathfrak{p}$ of length t from i to j, i.e. sequences of arcs $\mathfrak{p} = (e_1 f_1, \ldots, e_t f_t)$ with $e_k \in E_{i_{k-1}}^+$, $f_k \in E_{i_k}^-$ for some sequence $i_k \in V$ such that $i_0 = i$ and $i_t = j$, and where the *weight* is defined by

$$\mathbf{w}(\mathfrak{p}) \quad := \quad \frac{1}{d_{i_0}^+ \cdots d_{i_{t-1}}^+}.$$

As long as G is strongly connected (in the sense that for every $i, j \in V$ there exists a directed path from i to j), the classical theory guarantees that there is a unique probability measure $\pi_\star$ on V which is invariant, i.e. $\pi_\star P = \pi_\star$ (see, e.g., the book [27]). From every starting state $i \in V$, one may then consider the *total-variation distance from equilibrium at time t*:

$$\mathcal{D}_i(t) \quad := \quad \|P^t(i,\cdot) - \pi_\star\|_{\mathrm{TV}} \quad = \quad \frac{1}{2} \sum_{j \in V} \left|P^t(i,j) - \pi_\star(j)\right| \quad \in \quad [0,1]. \tag{2}$$

The aim of this paper is to investigate the profile of the decreasing functions $t \mapsto \mathcal{D}_i(t), i \in V$ under the *configuration model*, i.e. when the environment ω is chosen uniformly at random from the $m!$ possible choices. This turns G, P, $\pi_\star$ and the $(\mathcal{D}_i)_{i \in V}$ into random objects, parametrized by the degrees $(d_i^\pm)_{1 \leq i \leq n}$. In order to study large-size asymptotics, we let all quantities implicitly depend on n and consider the limit as $n \to \infty$. We restrict our attention to the *sparse regime*, where

$$\delta := \min_{1 \leq i \leq n} d_i^\pm \geq 2 \quad \text{and} \quad \Delta := \max_{1 \leq i \leq n} d_i^\pm = \mathcal{O}(1). \tag{3}$$

Note in particular that $m = \Theta(n)$. The requirement on δ guarantees that G is strongly connected with high probability (see, e.g., [13]). In particular, the equilibrium measure $\pi_\star$ is unique with high

probability. Here and below, with high probability (w.h.p.) means with probability tending to 1 as $n \to \infty$. The mixing time of the walk turns out to be determined by a simple statistics, namely the mean logarithmic out-degree of the end-point of a uniformly chosen head. More precisely, define

$$\mu := \frac{1}{m} \sum_{i=1}^{n} d_i^- \ln d_i^+ \quad \text{and} \quad t_\star := \frac{\ln n}{\mu}.$$

Our first result is that $t_\star$ steps are necessary and sufficient for the walk to mix, regardless of the initial vertex. More precisely, as the number of iterations t approaches $t_\star$, the distance from equilibrium undergoes a sharp transition visible on Figure 1 and known as a *cutoff phenomenon*.

Theorem 1 (Cutoff at time $t_\star$). *For $t = \lambda t_\star + o(t_\star)$ with fixed $\lambda \geq 0$, we have*

$$\lambda < 1 \quad \Longrightarrow \quad \min_{i \in V} \mathcal{D}_i(t) \xrightarrow[n \to \infty]{\mathbf{P}} 1,$$

$$\lambda > 1 \quad \Longrightarrow \quad \max_{i \in V} \mathcal{D}_i(t) \xrightarrow[n \to \infty]{\mathbf{P}} 0,$$

where $\xrightarrow{\mathbf{P}}$ denotes convergence in probability.

In view of Theorem 1, it is tempting to "zoom in" around the cutoff point $t_\star$ until the details of the abrupt transition from 1 to 0 become visible. The appropriate window-width turns out to be

$$w_\star := \frac{\sigma \sqrt{\ln n}}{\mu^{3/2}}, \quad \text{where} \quad \sigma^2 := \frac{1}{m} \sum_{i=1}^{n} d_i^- \left(\ln d_i^+ - \mu \right)^2.$$

Remarkably, the graph of the function $t \mapsto \mathcal{D}_i(t)$ inside this window approaches a universal shape, independent of the initial position i and the precise degrees: the gaussian tail function.

Theorem 2 (Inside the cutoff window). *Assume that the variance σ^2 is asymptotically non-degenerate in the following weak sense:*

$$\sigma^2 \gg \frac{(\ln \ln n)^2}{\ln n}. \tag{4}$$

Then, for $t = t_\star + \lambda w_\star + o(w_\star)$ with $\lambda \in \mathbb{R}$ fixed, we have

$$\max_{i \in V} \left| \mathcal{D}_i(t) - \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-\frac{u^2}{2}} \, du \right| \xrightarrow[n \to \infty]{\mathbf{P}} 0.$$

In contrast, our third main result asserts that the mixing time is reduced from $t_\star$ to constantly many steps when starting from a more spread-out distribution, such as the *in-degree distribution*:

$$\pi_0(i) := \frac{d_i^-}{m}, \qquad i \in V. \tag{5}$$

For future reference, the *out-degree distribution* is naturally defined by changing the $-$ to $+$ above.

3

Theorem 3 (Exponential convergence from a uniform head). *Fix $t \in \mathbb{N}$ and set $\pi_t := \pi_0 P^t$. Then,*

$$4\|\pi_t - \pi_\star\|_{\mathrm{TV}}^2 \quad \le \quad \frac{n(\gamma - 1)}{m(1 - \varrho)} \varrho^t + o_{\mathbf{P}}(1),$$

where $o_{\mathbf{P}}(1)$ denotes a term that tends to 0 in probability as $n \to \infty$, and where ϱ, γ are defined by

$$\varrho \ := \ \frac{1}{m} \sum_{i \in V} \frac{d_i^-}{d_i^+} \quad and \quad \gamma \ := \ \frac{1}{m} \sum_{i=1}^n \frac{(d_i^-)^2}{d_i^+}.$$

Note that $\varrho \le \frac{1}{\delta}$ and that the constant in front of ϱ^t is less than Δ. Thus, the convergence to equilibrium occurs exponentially fast, uniformly in n. Note also that $\gamma \ge 1$ by the Cauchy-Schwarz inequality, with equality if and only if $d_i^+ = d_i^-$ for all $i \in V$. In this very special case, it is not hard to see that $\pi_0 P = \pi_0$: the equilibrium measure $\pi_\star$ is nothing but the in-degree distribution π_0 itself, in agreement with the fact that the upper-bound given by Theorem 3 vanishes.

One surprising implication of Theorem 3 is that the equilibrium mass $\pi_\star(i)$ assigned to a vertex $i \in V$ is essentially determined by the (backward) local neighbourhood of i only. By combining this with a simple branching process approximation (see Section 7), we obtain rather precise asymptotics for the equilibrium measure. Specifically, our last main result asserts that the empirical distribution of the numbers $(n\pi_\star(i))_{i \in V}$ (as depicted in Figure 1) concentrates around a deterministic probability measure $\mathcal{L}$, explicitly identified as the law of the random variable

$$M_\star \ = \ \frac{n}{m} \sum_{k=1}^{d_{\mathcal{I}}^-} Z_k, \tag{6}$$

where $\mathcal{I}$ is uniformly distributed on V and independent of the $(Z_k)_{k \ge 1}$. The latter are i.i.d. mean-one random variables with law determined by the distributional fixed-point equation

$$Z_1 \ \overset{d}{=} \ \frac{1}{d_{\mathcal{J}}^+} \sum_{k=1}^{d_{\mathcal{J}}^-} Z_k, \tag{7}$$

where $\mathcal{J}$ has the out-degree distribution and is independent of $(Z_k)_{k \ge 1}$. This recursive distributional equation has been extensively studied, Z_1 being a special case of a random variable stable by weighted sum. Such self-similar variables appear notably in connections with Mandelbrot's multiplicative cascades and branching random walks. A more general version of (7) has been studied by Rösler [37], Liu [29, 28, 30, 31] and Barral [4, 5]. Among others, these references provide detailed results concerning the uniqueness of the solution Z_1, its left and right tails, its positive and negative moments, its support, and even its absolute continuity w.r.t. Lebesgue's measure.

To state our result, we recall that the $1-$Wasserstein (or Kantorovich-Rubinstein) distance between two probability measures $\mathcal{L}_1, \mathcal{L}_2$ on $\mathbb{R}$ (see, e.g., [39][Chapter 6]) is defined as

$$\mathcal{W}(\mathcal{L}_1, \mathcal{L}_2) \ = \ \sup_f \left| \int_{\mathbb{R}} f \, d\mathcal{L}_1 - \int_{\mathbb{R}} f \, d\mathcal{L}_2 \right|,$$

4

where the supremum runs over all $f \colon \mathbb{R} \to \mathbb{R}$ satisfying $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$.

Theorem 4 (Structure of the equilibrium measure). *Let $\mathcal{L}$ be the law of $M_\star$. Then,*

$$\mathcal{W}\left(\frac{1}{n}\sum_{i=1}^n \delta_{n\pi_\star(i)}, \mathcal{L}\right) \xrightarrow[n\to\infty]{\mathbf{P}} 0.$$

Remark 1 (Dependence on n). *Note that the deterministic measure $\mathcal{L}$ actually depends on n, as did the quantities $\mu, \sigma^2, \gamma, \varrho$ appearing in the above theorems: we have chosen to express all our approximations directly in terms of the true degree sequence $(d_i^{\pm})_{1 \leq i \leq n}$ (which we view as our input parameter), rather than loosing in generality by artificially assuming the weak convergence of the empirical degree distribution $\frac{1}{n}\sum_{i=1}^n \delta_{(d_i^-, d_i^+)}$ to some $n-$independent limit. Of course, any external assumption on the $n \to \infty$ behaviour of the degrees can then be "projected" onto the $n-$dependent constants provided by our results to yield bona fide convergence results, if needed.*

2 Related work

The phase transition described in Theorem 1 is an instance of the celebrated *cutoff phenomenon*, first singled out in the early 1980's in the context of card shuffling by Diaconis, Shahshahani and Aldous [22, 2, 3]. This remarkable discontinuity in the convergence to equilibrium of an ergodic Markov chain has since then been identified in a variety of contexts, ranging from random walks on groups to interacting particle systems. We refer the reader to [20, 38, 10] for more details. Perhaps surprisingly, the emergence of the gaussian shape inside the cutoff window in our Theorem 2 is not isolated: the very same feature has been observed in a few unrelated models, such as the random walk on the $n-$dimensional hypercube [21], or the simple-exclusion process on the circle [26].

Motivated by applications to real-world networks (see, e.g., the survey by Cooper [12] and the references therein), the mixing properties of random walks on large but finite random graphs have recently become the subject of many investigations. The attention has been mostly restricted to the undirected setting, where the in-degree distribution is reversible. In particular, Frieze and Cooper have studied the *cover time* (i.e., the expected time needed for the chain to visit all states) of various random graphs [14, 15, 16, 17], and analyzed the precise component structures induced by the walk on random regular graphs [19]. Bounds for the mixing time on the largest component of the popular Erdős–Renyi model have also been obtained by various authors, in both the critical and super-critical connectivity regime [35, 7, 24, 23].

More directly related to our work is the inspiring paper [33] by Lubetzky and Sly, which establishes the cutoff phenomenon and determines its precise window and shape for the simple and the non-backtracking random walks on random regular graphs. The results therein were very recently

generalized to all non-bipartite regular Ramanujan graphs by Lubetzky and Peres [32], thereby confirming a long-standing conjecture of Peres [36]. In [8], Berestycki, Lubetzky, Peres and Sly establish the cutoff phenomenon on the Erdős–Renyi model and on the more general configuration model, for both the simple and non-backtracking random walks. The latter case was simultaneously and independently addressed by Ben-Hamou and Salez [6], who additionally determine the precise second-order behaviour inside the cutoff window.

In contrast, very little is known about random walks on random directed graphs. The failure of the crucial *reversibility* property makes many of the ingredients used in the above works unavailable. Even understanding the equilibrium measure constitutes an important theoretical challenge, with applications to link-based ranking in large databases (see, e.g., [11] and the references therein). In [18], Cooper and Frieze consider the random digraph on n vertices formed by independently placing an arc between every pair of vertices with probability $p = \frac{d \ln n}{n}$, where $d > 1$ is fixed while $n \to \infty$. In this regime, they prove that the equilibrium measure is asymptotically close to the in-degree distribution. The recent work [1] by Addario-Berry, Balle and Perarnau provides precise estimates on the extrema of the equilibrium measure in a sparse random digraph where all out-degrees are equal. To the best of our knowledge, the present paper provides the first proof of the cutoff phenomenon in the non-reversible setting of random directed graphs.

3 Proof outline and main ingredients

3.1 Dealing with an unknown equilibrium measure

One difficulty in controlling the distance to equilibrium is that $\pi_\star$ is not known a priori. We will have to work instead with the proxy $\pi_h := \pi_0 P^h$, where π_0 is the in-degree distribution and

$$h \quad := \quad \left\lfloor \frac{\ln n}{10 \ln \Delta} \right\rfloor. \tag{8}$$

Establishing Theorems 1 and 2 with $\widetilde{\mathcal{D}}_i(t) := \|P^t(i, \cdot) - \pi_h\|_{\mathrm{TV}}$ instead of $\mathcal{D}_i(t)$ is actually sufficient. Indeed, it ensures in particular that for (say) $t \geq 2t_\star$,

$$\max_{i \in V} \|P^t(i, \cdot) - \pi_h\|_{\mathrm{TV}} \quad \xrightarrow[n \to \infty]{\mathbf{P}} \quad 0.$$

Now, by convexity, this maximum automatically extends to all initial distributions on V. In particular, one may start from the invariant measure $\pi_\star$ itself: since $\pi_\star P^t = \pi_\star$, we obtain

$$\|\pi_\star - \pi_h\|_{\mathrm{TV}} \quad \xrightarrow[n \to \infty]{\mathbf{P}} \quad 0. \tag{9}$$

By the triangle inequality, we finally deduce that

$$\sup_{i \in V, t \in \mathbb{N}} \left| \widetilde{\mathcal{D}}_i(t) - \mathcal{D}_i(t) \right| \quad \xrightarrow[n \to \infty]{\mathbf{P}} \quad 0,$$

so that the conclusions obtained for $\widetilde{\mathcal{D}}_i$ in Theorems 1 and 2 are automatically transferred to $\mathcal{D}_i$. In all proofs below, we will thus replace $\pi_\star$ with π_h and $\mathcal{D}_i$ with $\widetilde{\mathcal{D}}_i$, without explicit notice.

3.2 Sequential generation and tree-like structure

An elementary yet crucial observation about the uniform environment ω is that it can be generated sequentially, starting with all heads and tails unmatched, by repeating m times the following steps:

1. an unmatched tail e is selected according to some priority rule;

2. an unmatched head f is chosen uniformly at random;

3. e is matched with f to form the arc ef, that is, $\omega(e) := f$.

The resulting bijection ω is uniform, regardless of which particular priority rule is used. In the sequel, we shall intensively exploit this degree of freedom to simplify the analysis of the environment. In this respect, the following observation will prove useful. Let us say that a *collision* occurs whenever a head gets chosen, whose end-point i was already *alive* in the sense that some tail $e \in E_i^+$ or head $f \in E_i^-$ had previously been chosen. Since less than $2k$ vertices are alive when the kth head gets chosen, less than $2\Delta k$ of the $m - k + 1$ possible choices can result in a collision. Thus, the conditional chance that the kth arc causes a collision, given the past, is less than $\frac{2\Delta k}{m-k+1}$:

Lemma 5 (Collisions are rare). *Let $1 \leq k \leq m$. The number Z_k of collisions caused by the first k arcs is stochastically dominated by a $Binomial\left(k, \frac{2\Delta k}{m-k+1}\right)$ random variable. In particular,*

$$\mathbf{P}\left(Z_k \geq 1\right) \ \leq \ \frac{2\Delta k^2}{m-k+1} \quad and \quad \mathbf{P}\left(Z_k \geq 2\right) \ \leq \ \frac{2\Delta^2 k^4}{(m-k+1)^2}.$$

Here is one application: the *forward ball* of radius t around $i \in V$ is the subgraph $\mathcal{B}^+(i,t) \subseteq G$ induced by the directed paths of length t from i. It can be sequentially generated by giving priority to those unmatched tails e that currently lie at minimal distance from i, until this minimal distance exceeds t. At most $k = \Delta + \cdots + \Delta^t$ edges are formed by then, and each collision corresponds to the formation of a transverse arc violating the directed-tree structure of $\mathcal{B}^+(i,t)$. Choosing $t = 2h$ with h defined in (8) ensures that $\mathbf{P}\left(Z_k \geq 2\right) = o\left(\frac{1}{n}\right)$, uniformly in $i \in V$. We may thus take a union-bound and conclude that with high probability, G is locally tree-like in the following sense:

$$\forall i \in V, \quad \mathcal{B}^+(i, 2h) \text{ is either a directed tree, or a directed tree with an extra arc.}$$

Combining this with the fact that all out-degrees are at least 2, it is not difficult to deduce the following result. Let $V_\star$ denote the set of vertices $i \in V$ such that $\mathcal{B}^+(i, h)$ is a directed tree.

7

Proposition 6. *With high probability,*

$$\forall i \in V, \ \forall \ell \in \mathbb{N}, \qquad P^\ell(i, V \setminus V_\star) \ \leq \ 2^{-\ell \wedge h}.$$

In other words, $V_\star$ is typically attained in constant time by the random walk, from any initial position. This will be very helpful, as the walk is much easier to control when starting from $V_\star$.

Remark 2 (Sampling with or without replacement)**.** *Instead of choosing heads uniformly among the remaining unmatched ones, we may choose uniformly from all heads, and retry if the chosen head was already matched. The chance that this occurs within the first k steps is $p = 1 - \prod_{i=0}^{k-1} \frac{m-i}{m}$. This creates a coupling between the first k chosen heads and an i.i.d. sample from the uniform distribution on all heads, under which the two sequences coincide with probability $1 - p$. In the analysis of the sequential process, we may thus replace the former by the latter at a total-variation cost $p \leq k^2/m$. We may even stop the coupling as soon as a head is sampled, whose end-point was already alive: this ensures the absence of collision, while multiplying the total-variation cost p by a factor at most Δ. Note that the end-point of a uniform head has the in-degree distribution.*

3.3 Typical path weights

At a high level, the cutoff phenomenon around time $t_\star = \frac{\ln n}{\mu}$ described in Theorem 1 can be understood as the consequence of the following two key principles:

(i) Trajectories whose weight exceeds $\frac{1}{n}$ constitute the essential obstruction to mixing, in the precise sense that their total weight is roughly equal to the distance to equilibrium.

(ii) From the point-of-view of the walk, most trajectories of length t have weight $e^{-\mu t + \mathcal{O}(\sqrt{t})}$.

Thus, the essence of the cutoff phenomenon for the random walk on G lies in a sharp concentration phenomenon for the weight of the paths seen by the walk. To formalize this idea, let

$$\mathcal{Q}_{i,t}(\theta) \ := \ \sum_{j \in V} \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}) \mathbf{1}_{\mathbf{w}(\mathfrak{p}) > \theta}$$

denote the quenched probability that a random walk of length t starting at i follow a path whose weight exceeds θ. The above two claims can then be given the following precise meanings.

Proposition 7 (High-weight paths determine the distance to equilibrium)**.** *For any $t = t(n)$,*

$$\min_{i \in V} \mathcal{D}_i(t) \ \geq \ \min_{i \in V} \mathcal{Q}_{i,t}\left(\frac{\ln^3 n}{n}\right) - o_{\mathbf{P}}(1)$$

$$\max_{i \in V} \mathcal{D}_i(t) \ \leq \ \max_{i \in V} \mathcal{Q}_{i,t}\left(\frac{1}{n \ln^3 n}\right) + o_{\mathbf{P}}(1).$$

Proposition 8 (Most paths of length t have weight $e^{-\mu t + \mathcal{O}(\sqrt{t})}$). *Assume that $t = \Theta(\ln n)$, and let θ depend arbitrarily on n.*

1. *If $\frac{\mu t + \ln \theta}{\sqrt{t}} \to +\infty$ as $n \to \infty$, then*

$$\max_{i \in V} \mathcal{Q}_{i,t}(\theta) \xrightarrow[n \to \infty]{\mathbf{P}} 0.$$

2. *If $\frac{\mu t + \ln \theta}{\sqrt{t}} \to -\infty$ as $n \to \infty$, then*

$$\min_{i \in V} \mathcal{Q}_{i,t}(\theta) \xrightarrow[n \to \infty]{\mathbf{P}} 1.$$

3. *If $\frac{\mu t + \ln \theta}{\sigma \sqrt{t}} \to \lambda \in \mathbb{R}$ as $n \to \infty$ and if assumption (4) holds, then*

$$\max_{i \in V} \left| \mathcal{Q}_{i,t}(\theta) - \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-\frac{u^2}{2}} \, \mathrm{du} \right| \xrightarrow[n \to \infty]{\mathbf{P}} 0.$$

Those two results clearly imply Theorems 1 and 2, and their proofs occupy much of the paper: the lower and upper bounds in Proposition 7 are respectively established in sections 4 and 6, while Proposition 8 is proved in section 5. Once (9) is established, the branching process approximation for G developed in section 7 quickly leads to the proof of Theorems 3 and 4 in sections 8 and 9.

4 Proof of the lower-bound in Proposition 7

In this section, we prove the easy half of Proposition 7: the lower-bound. Fix the environment ω, a probability measure π on V, $t \in \mathbb{N}$, $\theta \in (0,1)$ and $i, j \in V$. Consider the inequality

$$P^t(i,j) \geq \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}) \mathbf{1}_{\mathbf{w}(\mathfrak{p}) \leq \theta}. \tag{10}$$

If equality holds in (10), then clearly

$$\pi(j) - \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}) \mathbf{1}_{\mathbf{w}(\mathfrak{p}) \leq \theta} \leq \left[\pi(j) - P^t(i,j) \right]_+,$$

where $[x]_+ := \max(x, 0)$ is the positive part of x. On the other-hand, if the inequality (10) is strict, then there must exist $\mathfrak{p} \in \mathcal{P}_{ij}^t$ such that $\mathbf{w}(\mathfrak{p}) > \theta$, implying that $P^t(i,j) > \theta$, and hence that

$$\pi(j) - \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}) \mathbf{1}_{\mathbf{w}(\mathfrak{p}) \leq \theta} \leq \pi(j) \mathbf{1}_{P^t(i,j) > \theta}.$$

Thus, in either case,

$$\pi(j) - \sum_{\mathfrak{p} \in \mathcal{P}_{ij}^t} \mathbf{w}(\mathfrak{p}) \mathbf{1}_{\mathbf{w}(\mathfrak{p}) \leq \theta} \leq \left[\pi(j) - P^t(i,j) \right]_+ + \pi(j) \mathbf{1}_{P^t(i,j) > \theta}.$$

9

Summing over all $j \in V$ and invoking the Cauchy–Schwarz and Markov inequalities, we get

$$\mathcal{Q}_{i,t}(\theta) \leq \|\pi - P^t(i, \cdot)\|_{\mathrm{TV}} + \sqrt{\frac{1}{\theta} \sum_{j \in V} \pi^2(j)}.$$

We now specialize to our random environment, with $\theta = \frac{\ln^3 n}{n}$ and $\pi = \pi_h$. To conclude the proof, it remains to verify that the square-root term is $o_{\mathbf{P}}(1)$. We will in fact show the stronger

$$\mathbf{E}\left[\sum_{j \in V} \pi_h^2(j)\right] \ll \theta.$$

Since $\pi_h = \pi_0 P^h$, the left-hand-side may be interpreted as $\mathbf{P}(X_h = Y_h)$, where conditionally on the environment, $(X_k)_{0 \leq k \leq h}$ and $(Y_k)_{0 \leq k \leq h}$ are two independent random walks starting from the in-degree distribution π_0, and where the average is taken over both the walks and the environment. To evaluate this annealed probability, we generate the walks together with the environment, forming arcs along the way, as we need them. Initially, all tails and heads are unmatched, and X_0 is chosen according to the in-degree distribution. Then, inductively, at every step $1 \leq k \leq h$:

1. A tail e of the vertex X_{k-1} is chosen uniformly at random.

2. If e is unmatched, it gets matched to a uniformly chosen unmatched head f, i.e. $\omega(e) := f$.

3. In either case, $\omega(e)$ is now well-defined, and we let X_k be its end-point.

Once $(X_k)_{0 \leq k \leq h}$ has been generated, we proceed similarly with $(Y_k)_{0 \leq k \leq h}$. Note that at most $2h$ arcs are formed during this process. For the event $\{X_h = Y_h\}$ to be realized, either a collision must occur (this has probability at most $\frac{8\Delta h^2}{m - 2h}$ by Lemma 5), or $\{Y_0 = X_0\}$ must occur (this has probability at most $\frac{\Delta}{m}$). We deduce that

$$\mathbf{P}(X_h = Y_h) \leq \frac{8\Delta(h+1)^2}{m - 2h}.$$

The right-hand side is $\mathcal{O}(\frac{\ln^2 n}{n})$. In view of our choice of θ, this concludes the proof.

5 Proof of Proposition 8

Recall that Proposition 8 controls, uniformly in $i \in V$, the quenched probability $\mathcal{Q}_{i,t}(\theta)$ that a random walk of length t starting at i follows a path whose weight exceeds θ. We first restrict our attention to vertices i in $V_\star$, and replace $\mathcal{Q}_{i,t}(\theta)$ with the more stable spatial average

$$\overline{\mathcal{Q}}_{i,t}(\theta) := \sum_{j \in V} P^\ell(i, j) \mathcal{Q}_{j,t}(\theta), \qquad \text{where} \qquad \ell := \lfloor 3 \ln \ln n \rfloor.$$

10

Let $(D_k^+)_{k \geq 1}$ denote i.i.d. copies of $d_{\mathcal{J}}^+$ where $\mathcal{J}$ follows the in-degree distribution on V. More explicitly, $\mathbf{P}\left(D_k^+ = d\right) = \frac{1}{m}\sum_{i=1}^n d_i^- \mathbf{1}_{(d_i^+ = d)}$. For $\theta \in [0,1]$, define

$$q_t(\theta) \quad := \quad \mathbf{P}\left(\prod_{k=1}^t \frac{1}{D_k^+} > \theta\right).$$

Lemma 9. *For $t = \Theta(\ln n)$ and θ depending arbitrarily on n,*

$$\max_{i \in V_\star}\left|\overline{\mathcal{Q}}_{i,t}(\theta) - q_t(\theta)\right| \quad \xrightarrow[n \to \infty]{\mathbf{P}} \quad 0.$$

Proof. Given the environment ω, consider $r = \lfloor \ln^2 n \rfloor$ independent walks of length $\ell + t$ starting at $i \in V$. Let A_i be the event that their trajectories up to time ℓ form a tree and that their trajectories after time ℓ all have weights $> \theta$. Then $\mathbf{1}_{i \in V_\star}\left(\overline{\mathcal{Q}}_{i,t}(\theta)\right)^r$ is clearly a lower-bound for the quenched probability of A_i. Averaging over ω and using Markov's inequality, we get for $\varepsilon > 0$,

$$\mathbf{P}\left(i \in V_\star, \overline{\mathcal{Q}}_{i,t}(\theta) \geq q_t(\theta) + \varepsilon\right) \quad \leq \quad \frac{\mathbf{P}(A_i)}{(q_t(\theta) + \varepsilon)^r}. \tag{11}$$

To evaluate the annealed probability $\mathbf{P}(A_i)$, we may generate the r walks one after the other together with the environment, forming arcs only when they are traversed by the walks, as we did in the previous section. Given that the first $k - 1$ walks do satisfy the requirement, the conditional chance that the kth walk also does is at most $q_t(\theta) + o(1)$ uniformly in $i \in V$ and $1 \leq k < r$. Indeed,

- either the kth walk attains length ℓ before reaching an unmatched tail: thank to the tree structure, there are at most $k - 1 < r$ possible trajectories to follow, and each has weight at most $2^{-\ell}$ by our assumption on the minimum degrees. Thus, the conditional chance of this scenario is less than $r 2^{-\ell} = o(1)$.

- or the walk has reached an unmatched tail by time ℓ: as explained in remark 2, the remainder of the trajectory after the first unmatched tail can be coupled with an i.i.d. sample from the in-degree distribution on V at a total-variation cost less than $\frac{\Delta r (t + \ell)^2}{m} = o(1)$. Thus, the conditional chance that the walk meet the requirement in that case is at most $q_t(\theta) + o(1)$.

This proves that $\mathbf{P}(A_i) \leq (q_t(\theta) + o(1))^r$, uniformly in $i \in V$. Inserting into (11) and summing over all possible $i \in V$, we obtain the first half of the claim:

$$\mathbf{P}\left(\exists i \in V_\star, \overline{\mathcal{Q}}_{i,t}(\theta) \geq q_t(\theta) + \varepsilon\right) \quad \xrightarrow[n \to \infty]{} \quad 0.$$

Replacing "$> \theta$" by "$\leq \theta$" in the definitions of A_i, $\overline{\mathcal{Q}}_{i,t}(\theta)$ and $q_t(\theta)$ yields the other half. $\qquad \square$

Proof of Proposition 8. Recalling the definitions of $\mathcal{Q}_{i,t}(\theta)$ and $\overline{\mathcal{Q}}_{i,t}(\theta)$, we have

$$\max_{i \in V} \mathcal{Q}_{i,t}(\theta) \;\; \leq \;\; \max_{i \in V_\star} \overline{\mathcal{Q}}_{i,t-2\ell}(\theta) + \max_{i \in V} P^\ell(i, V \setminus V_\star).$$

Now, observe that $\overline{\mathcal{Q}}_{i,t-s}(\theta) \leq \overline{\mathcal{Q}}_{i,t}(\theta\Delta^{-s})$ for all $s \leq t$. Choosing $s = 2\ell$, we obtain

$$\begin{aligned} \max_{i \in V} \mathcal{Q}_{i,t}(\theta) \;\; &\leq \;\; \max_{i \in V_\star} \overline{\mathcal{Q}}_{i,t}(\theta\Delta^{-2\ell}) + \max_{i \in V} P^\ell(i, V \setminus V_\star) \\ &\leq \;\; q_t(\theta\Delta^{-2\ell}) + o_{\mathbf{P}}(1), \end{aligned} \tag{12}$$

where in the second line we have used Lemma 9 and Proposition 6. Similarly, we may write

$$\mathcal{Q}_{i,t}(\theta) \;\; \geq \;\; P^\ell(i, V_\star) \min_{j \in V_\star} \overline{\mathcal{Q}}_{j,t-2\ell}(\theta\Delta^{2\ell}),$$

and then use $P^\ell(i, V_\star) = 1 - P^\ell(i, V \setminus V_\star)$ and the fact that $\overline{\mathcal{Q}}_{j,t}(\theta)$ is non-increasing in t to obtain

$$\begin{aligned} \min_{i \in V} \mathcal{Q}_{i,t}(\theta) \;\; &\geq \;\; \min_{i \in V_\star} \overline{\mathcal{Q}}_{i,t}(\theta\Delta^{2\ell}) - \max_{i \in V} P^\ell(i, V \setminus V_\star) \\ &\geq \;\; q_t(\theta\Delta^{2\ell}) - o_{\mathbf{P}}(1), \end{aligned} \tag{13}$$

where the last inequality follows again from Lemma 9 and Proposition 6. Recalling that $(\ln D_k^+)_{k \geq 0}$ are i.i.d. with mean μ and variance σ^2, we may now conclude as follows:

- If $\frac{\mu t + \ln \theta}{\sqrt{t}} \to +\infty$ then $q_t(\theta) \to 0$, by Chebychev's inequality.

- If $\frac{\mu t + \ln \theta}{\sqrt{t}} \to -\infty$ then $q_t(\theta) \to 1$, by Chebychev's inequality again.

- If $\frac{\mu t + \ln \theta}{\sigma\sqrt{t}} \to \lambda \in \mathbb{R}$ and (4) holds, then by Berry–Esseen's inequality (recall that the random variable $\ln D_k^+$ is uniformly bounded thanks to (3)),

$$q_t(\theta) \xrightarrow[n \to \infty]{} \frac{1}{\sqrt{2\pi}} \int_\lambda^\infty e^{-\frac{u^2}{2}} \, \mathrm{du}.$$

In each case, the assumption on θ is insensitive to multiplication by $\Delta^{\pm 2\ell}$: indeed, this only shifts $\ln \theta$ by $\mathcal{O}(\ln \ln n)$, which is negligible compared to $\sqrt{t}$, and even to $\sigma\sqrt{t}$ under the extra assumption (4). Consequently, the same three conclusions hold with θ replaced by $\theta\Delta^{\pm 2\ell}$, and inserting into (12) and (13) concludes the proof of Proposition 8. $\qquad \square$

6 Proof of the upper-bound in Proposition 7

6.1 General strategy

Fix the environment and consider the lower-bound $P_0^t(i, j) \leq P^t(i, j)$ obtained by restricting the sum (1) to certain *nice* paths, to be specified. Assume it is small enough so that for all $i, j \in V$,

$$P_0^t(i, j) \;\; \leq \;\; (1 + \varepsilon)\pi(j) + \frac{\varepsilon}{|V|}, \tag{14}$$

for a given $\varepsilon > 0$ and a given probability measure π on V. Then clearly,

$$
\begin{aligned}
\|\pi - P^t(i, \cdot)\|_{\mathrm{TV}} &= \sum_{j \in V} \left(\pi(j) - P^t(i,j) \right)_+ \\
&\leq \sum_{j \in V} \left(\pi(j)(1 + \varepsilon) + \frac{\varepsilon}{|V|} - P_0^t(i,j) \right)_+ = p(i) + 2\varepsilon,
\end{aligned}
$$

where $p(i) = 1 - P_0^t(i, V)$ is the quenched probability that the path followed by the random walk of length t starting at i is not *nice*. We use this strategy with $\pi = \pi_h$ and the following nice paths.

Definition 1. *A path $\mathfrak{p}$ of length t starting at i is* nice *if it satisfies the following conditions:*

1. $\mathbf{w}(\mathfrak{p}) \leq \frac{1}{n \ln^2 n}$;

2. *the first $t - h$ steps of $\mathfrak{p}$ are contained in a certain subtree $\mathcal{T}_i \subseteq G$, defined below;*

3. *the last h steps of $\mathfrak{p}$ form the only path of length at most h from its origin to destination.*

With this particular definition of nice paths, we will establish the following result.

Proposition 10. *Fix $\varepsilon > 0$ and assume that $t = t_\star + o(t_\star)$. Then with high probability,*

1. *every pair $(i, j) \in V^2$ satisfies (14);*

2. *every $i \in V_\star$ satisfies $p(i) \leq \mathcal{Q}_{i,t}\left(\frac{1}{n \ln^2 n} \right) + \varepsilon$.*

As a corollary, we immediately deduce that for $t = t_\star + o(t_\star)$,

$$
\max_{i \in V_\star} \mathcal{D}_i(t) \leq \max_{i \in V_\star} \mathcal{Q}_{i,t}\left(\frac{1}{n \ln^2 n} \right) + o_{\mathbf{P}}(1). \tag{15}
$$

This statement easily extends to arbitrary $t = t(n)$ as follows. For, say, $t \leq t_\star - t_\star^{\frac{2}{3}}$, we may use $\mathcal{D}_i(t) \leq 1$, part 2 of Proposition 8 and the fact that $\mathcal{Q}_{i,t}(\theta)$ is non-increasing in t. For $t \geq t_\star + t_\star^{\frac{2}{3}}$, we may use part 1 of Proposition 8 and the fact that $\mathcal{D}_i(t)$ is non-increasing in t. Finally, to extend the maximum on the left to all vertices, we take $\ell = \lfloor \frac{\ln \ln n}{\ln \Delta} \rfloor$ and observe that

$$
\begin{aligned}
\max_{i \in V} \mathcal{D}_i(t) &\leq \max_{i \in V_\star} \mathcal{D}_i(t - \ell) + o_{\mathbf{P}}(1) \\
&\leq \max_{i \in V_\star} \mathcal{Q}_{i,t-\ell}\left(\frac{1}{n \ln^2 n} \right) + o_{\mathbf{P}}(1) \\
&\leq \max_{i \in V_\star} \mathcal{Q}_{i,t}\left(\frac{1}{n \ln^3 n} \right) + o_{\mathbf{P}}(1),
\end{aligned}
$$

where we have successively used Proposition 6, (15), and the fact that $\mathcal{Q}_{i,t-\ell}(\theta) \leq \mathcal{Q}_{i,t}\left(\frac{\theta}{\Delta^\ell} \right)$.

6.2 Construction of the subtree $\mathcal{T}_i$ and proof of the first part in Proposition 10

The following procedure generates the environment ω locally around a given origin $i \in V$ and extracts from it a certain subtree $\mathcal{T} = \mathcal{T}_i(\omega)$. Throughout the process, we let $\partial_+\mathcal{T}$ (resp. $\partial_-\mathcal{T}$) denote the set of unmatched tails (resp. heads) whose vertex belongs to $\mathcal{T}$. Initially, all tails and heads are unmatched and $\mathcal{T}$ is reduced to its root, i. We then iterate the following steps:

1. A tail $e \in \partial_+\mathcal{T}$ is selected according to some rule, see below.

2. e is matched to a uniformly chosen unmatched head f, forming an arc ef.

3. If f was not in $\partial_-\mathcal{T}$, then the arc ef and its end-vertex are added to the subtree $\mathcal{T}$.

The condition in step 3 ensures that $\mathcal{T}$ remains a directed tree: any $e \in \partial_+\mathcal{T}$ is accessible from the root by a unique directed path, and we define the *height* $\mathbf{h}(e)$ as the number of vertices along that path (including the vertex of e), and the *weight* $\mathbf{w}(e)$ as the inverse product of vertex out-degrees along that path (including the vertex of e). Our rule for step 1 consists in selecting a tail with maximal weight[1] among all $e \in \partial_+\mathcal{T}$ with height $\mathbf{h}(e) < t - h$ and weight $\mathbf{w}(e) > w_{\min}$, where

$$w_{\min} \quad := \quad \frac{\ln n}{n}.$$

The procedure stops when there is no such tail, which occurs after a random number κ of pairings. The only role of the parameter $w_{\min}$ is to control κ. Specifically, it will follow from Lemma 13 that

$$\kappa \quad \le \quad \frac{2}{w_{\min}} \quad = \quad o(n). \tag{16}$$

At the end of the procedure, we let $\mathcal{E}$ be the set of tails $e \in \partial_+\mathcal{T}$ such that $\mathbf{h}(e) = t - h$. Note that

$$\sum_{e \in \mathcal{E}} \mathbf{w}(e) \quad \le \quad 1. \tag{17}$$

We next sequentially generate the *backward ball of radius h* around a given destination $j \in V$ as we did for the forward ball in subsection 3.2, but with all orientations reversed. This creates an additional random number τ of arcs, satisfying the crude bound

$$\tau \quad \le \quad \Delta + \cdots + \Delta^h \quad = \quad o(n). \tag{18}$$

We may now consider the set $\mathcal{F}$ of heads f from the end-point of which the shortest path to j is unique and has length h. Writing $\mathbf{w}(f)$ for the weight of that path, we have by definition of π_h,

$$\pi_h(j) \quad \ge \quad \frac{1}{m} \sum_{f \in \mathcal{F}} \mathbf{w}(f). \tag{19}$$

---

[1]using an arbitrary deterministic ordering of the set of all tails to break ties.

Finally, we complete the environment by matching the $m - \kappa - \tau$ remaining unmatched heads to the $m - \kappa - \tau$ remaining unmatched heads uniformly at random. Recalling definition 1,

$$P_0^t(i,j) \;=\; \sum_{e \in \mathcal{E}} \sum_{f \in \mathcal{F}} \mathbf{w}(e)\mathbf{w}(f) \mathbf{1}_{\mathbf{w}(e)\mathbf{w}(f) \leq \frac{1}{n \ln^2 n}} \mathbf{1}_{\omega(e)=f}.$$

By construction, any arc ef contributing to this sum must be formed during the completion stage. Conditionally on the $\sigma-$field $\mathcal{G}$ generated by the $\kappa + \tau$ arcs formed before that, the random variable $P_0^t(i,j)$ may thus be regarded as the cost of a uniform bijection, and we can exploit the following concentration result due to Chatterjee [9].

Theorem 11 ([9], Proposition 1.1). *Fix two finite sets E, F with $|E| = |F|$ and $c \colon E \times F \to \mathbb{R}_+$. Consider the random variable $Z := \sum_{e \in E} c(e, \omega(e))$ where $\omega \colon E \to F$ is a uniform bijection. Then,*

$$\mathbf{P}\left(Z - \mathbf{E}[Z] \geq \eta\right) \;\leq\; \exp\left(-\frac{\eta^2}{2\|c\|_\infty \left(2\mathbf{E}[Z] + \eta\right)}\right),$$

for every $\eta > 0$, where $\|c\|_\infty = \max_{(e,f) \in E \times F} c(e,f)$.

We apply this result to $Z = P_0^t(i,j)$, conditionally on $\mathcal{G}$. Note that in our case, $\|c\|_\infty \leq \frac{1}{n \ln^2 n}$ by construction. Moreover,

$$\mathbf{E}[Z|\mathcal{G}] \;\leq\; \frac{1}{m - \kappa - \tau}\left(\sum_{e \in \mathcal{E}} \mathbf{w}(e)\right)\left(\sum_{f \in \mathcal{F}} \mathbf{w}(f)\right) \;\leq\; (1 + o(1))\pi_h(j),$$

thanks to (16), (17), (18) and (19), with the $o(1)$ term being deterministic and independent of i, j. For n large enough, it is smaller than $\varepsilon/2$, and choosing $\eta := \frac{\varepsilon}{2}\mathbf{E}[Z|\mathcal{G}] + \frac{\varepsilon}{n}$ then yields

$$\max_{(i,j) \in V^2} \mathbf{P}\left(P_0^t(i,j) \geq (1 + \varepsilon)\pi_h(j) + \frac{\varepsilon}{n}\right) \;\leq\; \exp\left(-\frac{\eta^2 \ln^2 n}{2n(2\mathbf{E}[Z|\mathcal{G}] + \eta)}\right)$$

$$\leq\; \exp\left(-\frac{\varepsilon^2 \ln^2 n}{2(4 + \varepsilon)}\right).$$

We may thus take a union bound over all $(i,j) \in V^2$, establishing the first part of Proposition 10.

6.3   Proof of the second part of Proposition 10

It now remains to bound the quenched probability that the trajectory of the random walk of length t starting at $i \in V_\star$ is not nice. The first requirement in definition 1 fails with probability exactly $\mathcal{Q}_{i,t}\left(\frac{1}{n \ln^2 n}\right)$. A failure of the third requirement implies that the $(t - h)^{\text{th}}$ vertex is not in $V_\star$, which has probability $o_{\mathbf{P}}(1)$ uniformly in $i \in V$ thanks to Proposition 6. Regarding the second requirement, there are, by construction, only two ways of escaping from $\mathcal{T}_i$ before time $t - h$:

- Either the weight of the trajectory falls below $w_{\min} = \frac{\ln n}{n}$ before time $t - h$: the quenched probability of this is $1 - \mathcal{Q}_{i,t-h}(\frac{\ln n}{n})$, which is $o_{\mathbf{P}}(1)$ uniformly in i by Proposition 8.

- Or the walk traverses an arc that violated the tree structure in step 3 above. Proposition 12 below will show that the quenched probability of this is $o_{\mathbf{P}}(1)$ uniformly in $i \in V_\star$.

Consider again the construction of the subtree $\mathcal{T}$ described above. For $1 \leq k \leq \kappa$, we let $e_k f_k$ denote the kth formed arc, $\mathcal{T}^k$ the tree obtained after k arcs have been formed, and we consider the process $(W_k)_{k \geq 0}$ defined by $W_0 = 0$ and then inductively,

$$W_{k+1} \;=\; W_k + \mathbf{1}_{k < \kappa} \mathbf{1}_{f_{k+1} \in \partial_- \mathcal{T}^k} \mathbf{w}(e_{k+1}).$$

Thus, W_κ is the total weight of all tails that violated the tree structure in step 3 above. Since $W_h = 0$ for $i \in V_\star$, the following proposition is more than enough to complete the proof of Proposition 10.

Proposition 12. *For any fixed $\varepsilon > 0$, we have uniformly in $i \in V$,*

$$\mathbf{P}\left(W_\kappa \geq W_{\lfloor 2/\varepsilon \rfloor} + \varepsilon\right) \;=\; o\left(\frac{1}{n}\right).$$

The proof of this result will exploit the following auxiliary Lemma.

Lemma 13. *Amost-surely, for all $1 \leq k \leq \kappa$,*

$$\mathbf{w}(e_k) \;\leq\; \frac{2}{k+2}.$$

Since $\mathbf{w}(e_\kappa) > w_{\min}$ by our selection rule, this establishes in particular the announced bound (16).

Proof. It is enough to fix $w > 0$ and prove that whenever some $1 \leq k \leq \kappa$ satisfies

$$\mathbf{w}(e_k) \;\geq\; w, \tag{20}$$

we necessarily also have

$$\frac{2}{k+2} \;\geq\; w. \tag{21}$$

Let us assume that (20) holds. Since the condition (21) is monotone in k we may further assume that k is maximal with property (20). Thus, $\mathbf{w}(e_1) \geq \ldots \geq \mathbf{w}(e_k) \geq w > \mathbf{w}(e_{k+1}) \geq \cdots \geq \mathbf{w}(e_\kappa)$. The tree $\mathcal{T}^k$ may not contain all the k arcs $e_1 f_1, \ldots, e_k f_k$, because some of them may have caused a collision upon creation. Let us complete the tree by viewing the tails of those missing arcs as proper pending edges ending in a leaf node. Observe that in the resulting tree every non-leaf node i has exactly d_i^+ outgoing arcs, all having the same weight. In particular, this new tree has the following properties:

1. there are exactly k arcs;

2. every non-leaf node has at least two children;

3. every arc has weight at least w;

4. the weights of the arcs incident to the leaves sum up to 1.

A simple induction over the number of non-leaf nodes shows that any tree satisfying properties 1-2 has at least $\frac{k+2}{2}$ leaves. On the other hand, any tree satisfying properties 3-4 has at most $\frac{1}{w}$ leaves. Thus, (21) must hold. □

Proof of Proposition 12. Let $(\mathcal{G}_k)_k$ be the natural filtration associated with the construction of $\mathcal{T}$, i.e. $\mathcal{G}_k = \sigma(f_1, \ldots, f_k)$. Note that κ is a stopping time, that e_{k+1} is $\mathcal{G}_k$−measurable, and that the conditional law of f_{k+1} given $\mathcal{G}_k$ is uniform on the $m - k$ unmatched tails. Consequently,

$$
\mathbf{E}\left[W_{k+1} - W_k \middle| \mathcal{G}_k \right] = \mathbf{1}_{k<\kappa}\mathbf{w}(e_{k+1})\frac{|\partial_-\mathcal{T}^k|}{m-k},
$$

$$
\mathbf{E}\left[(W_{k+1} - W_k)^2 \middle| \mathcal{G}_k \right] = \mathbf{1}_{k<\kappa}\mathbf{w}(e_{k+1})^2\frac{|\partial_-\mathcal{T}^k|}{m-k}.
$$

Using Lemma 13, its corollary (16), and the crude bound $|\partial_-\mathcal{T}^k| \leq \Delta(k+1)$, we arrive at

$$
\sum_{k=0}^{\kappa-1} \mathbf{E}\left[W_{k+1} - W_k \middle| \mathcal{G}_k \right] \leq \frac{8\Delta/w_{\min}}{m - 2/w_{\min}} =: \alpha, \tag{22}
$$

$$
\sum_{k=0}^{\kappa-1} \mathbf{E}\left[(W_{k+1} - W_k)^2 \middle| \mathcal{G}_k \right] \leq \frac{4\Delta\ln(2e/w_{\min})}{m - 2/w_{\min}} =: \beta. \tag{23}
$$

where we have used $\sum_{k=0}^{\kappa-1}\frac{1}{k+3} \leq 1 + \ln\kappa$. In addition, Lemma 13 guarantees the a.-s. bound

$$
0 \leq W_{k+1} - W_k \leq \frac{2}{k+3}. \tag{24}
$$

In view of (22),(23),(24), the martingale version of Bennett's inequality due to Freedman [25, Theorem 1.6] ensures that for every $\varepsilon > \alpha, k \in \mathbb{N}$,

$$
\mathbf{P}\left(W_\kappa - W_k \geq \varepsilon \right) \leq \left(\frac{e(k+3)\beta}{2(\varepsilon - \alpha)} \right)^{\frac{(\varepsilon-\alpha)(k+3)}{2}}.
$$

Note that α and β do not depend on i and satisfy $\alpha = o(1)$ and $\beta = n^{-1+o(1)}$. Consequently, the right hand-side is $o\left(n^{-1}\right)$ as soon as $k + 3 > 2/\varepsilon$, and this concludes the proof. □

17

7 The martingale approximation

In this section, we show that $\mathcal{B}^-(\mathcal{I}, h)$, the backward ball of radius h around a uniform vertex $\mathcal{I}$, can be accurately described by the first h generations of a certain Galton-Watson tree $\mathcal{T}_\star$, allowing us to approximate the sequence $(n\pi_t(\mathcal{I}))_{0 \le t \le h}$ by a martingale defined directly on $\mathcal{T}_\star$. Specifically, let $\mathcal{T}_\star$ be the infinite random tree with marks in V obtained by the following branching process:

- Generation 0 consists of a single node (the root o), with mark uniformly distributed in V.

- Inductively, every node x in generation $t \in \mathbb{N}$ independently gives birth to exactly $d^-_{i(x)}$ children, where $i(x)$ denotes the mark of x. These new nodes form generation $t+1$, and their marks are chosen independently from the out-degree distribution on V.

Now, consider a node x at generation $t \in \mathbb{N}$, and let $(x_0, \dots, x_t)$ be the unique path from x to the root (thus, $x_0 = x$ and $x_t = o$). Define the weight of x as

$$\mathbf{w}(x) \quad := \quad \frac{nd^-_{i(x_0)}}{m} \prod_{k=0}^{t-1} \frac{1}{d^+_{i(x_k)}}.$$

In the case where $x = o$, we have $t = 0$ and the empty product is interpreted as equal to 1. Since $\mathbf{E}[d^-_{\mathcal{I}}] = 1$ when $\mathcal{I}$ is uniformly distributed over V, we have $\mathbf{E}[\mathbf{w}(o)] = 1$. For $t \in \mathbb{N}$, we now define M_t as the total weight of the tth generation $\mathcal{T}^t_\star$:

$$M_t \quad := \quad \sum_{x \in \mathcal{T}^t_\star} \mathbf{w}(x).$$

The law of the random variable M_t provides a good approximation to the law of $n\pi_t(\mathcal{I})$, where $\mathcal{I}$ denotes a uniformly chosen vertex, independent of the environment.

Proposition 14 (Branching process approximation)**.** *The total variation distance between the law of the random vector $(n\pi_t(\mathcal{I}))_{0 \le t \le h}$ and that of $(M_t)_{0 \le t \le h}$ is less than $\frac{\Delta^{2h+3}}{m}$.*

Proof. We may generate $\mathcal{B}^-(\mathcal{I}, h)$ sequentially as we did for the forward ball in subsection 3.2, with directions reversed. It is now tails that get uniformly chosen from the remaining unmatched ones. Building on remark 2, we may instead choose uniformly from all tails, matched or not, until a tail gets chosen whose end-point was already in the ball. The chance that this event occurs before the end of the procedure is less than $p = \Delta k^2/m$, where $k = \Delta + \cdots + \Delta^h \le \Delta^{h+1}$ is a crude upper-bound on the total number of steps. This creates a coupling between the ball $\mathcal{B}^-(\mathcal{I}, h)$ and the first h generations of the tree $\mathcal{T}_\star$, under which they coincide with probability more than $1 - p$. Moreover, on this event, $(M_t)_{0 \le t \le h}$ equals $(n\pi_t(\mathcal{I}))_{0 \le t \le h}$ by construction. $\square$

This connection to $(n\pi_t(\mathcal{I}))_{0 \le t \le h}$ motivates a deeper study of the process $(M_t)_{t \ge 0}$.

Proposition 15 (The martingale). $(M_t)_{t \geq 0}$ *is a martingale relative to the natural filtration of the branching process. The limit* $M_\star = \lim_{t \to \infty} M_t$ *exists almost-surely and in* L^2 *and for all* $t \in \mathbb{N}$,

$$\mathbf{E}\left[(M_\star - M_t)^2\right] \;\; = \;\; \frac{n(\gamma-1)\varrho^t}{m(1-\varrho)}.$$

Proof. Write $\to$ for the child to parent relation in $\mathcal{T}_\star$. By definition,

$$M_{t+1} - M_t \;\; = \;\; \sum_{x \in \mathcal{T}_\star^t} \frac{\mathbf{w}(x)}{d_{i(x)}^-} \sum_{y \to x} \left(\frac{d_{i(y)}^-}{d_{i(y)}^+} - 1\right).$$

Given $\mathcal{F}_t$, the random vector $(i(y))_{y \to x}$ consists of $d_{i(x)}^-$ i.i.d. samples from the out-degree distribution on V. Now, for a variable $\mathcal{J}$ with this distribution, we have

$$\mathbf{E}\left[\frac{d_{\mathcal{J}}^-}{d_{\mathcal{J}}^+}\right] \;=\; \sum_{i=1}^n \frac{d_i^+}{m} \frac{d_i^-}{d_i^+} \;=\; 1 \quad \text{and} \quad \mathbf{E}\left[\left(\frac{d_{\mathcal{J}}^-}{d_{\mathcal{J}}^+}\right)^2\right] \;=\; \sum_{i=1}^n \frac{d_i^+}{m} \left(\frac{d_i^-}{d_i^+}\right)^2 \;=\; \gamma.$$

Consequently, $\mathbf{E}\left[M_{t+1} - M_t \,|\, \mathcal{F}_t\right] = 0$ and

$$\mathbf{Var}\left[M_{t+1} - M_t \,|\, \mathcal{F}_t\right] \;\; = \;\; (\gamma-1) \sum_{x \in \mathcal{T}_\star^t} \frac{(\mathbf{w}(x))^2}{d_{i(x)}^-} =: \Sigma_t.$$

This shows that $(M_t)_{t \in \mathbb{N}}$ is a (non-negative) martingale, and that its almost-sure limit $M_\star$ satisfies

$$\mathbf{E}\left[(M_t - M_\star)^2\right] \;\; = \;\; \sum_{k=t}^\infty \mathbf{E}\left[\Sigma_k\right], \tag{25}$$

provided the right-hand side is finite. We now compute $\mathbf{E}[\Sigma_t]$ for all $t \in \mathbb{N}$. First note that

$$\Sigma_{t+1} \;\; = \;\; (\gamma-1) \sum_{x \in \mathcal{T}_\star^t} \left(\frac{\mathbf{w}(x)}{d_{i(x)}^-}\right)^2 \sum_{y \to x} \frac{d_{i(y)}^-}{\left(d_{i(y)}^+\right)^2}.$$

Now, recall that given $\mathcal{F}_t$, the variables $(i(y))_{y \to x}$ are $d_{i(x)}^-$ i.i.d. samples from the out-degree distribution on V, and observe that for a variable $\mathcal{J}$ with this distribution, we have

$$\mathbf{E}\left[\frac{d_{\mathcal{J}}^-}{\left(d_{\mathcal{J}}^+\right)^2}\right] \;\; = \;\; \sum_{i=1}^n \frac{d_i^+}{m} \frac{d_i^-}{(d_i^+)^2} \;\; = \;\; \varrho.$$

Consequently, $\mathbf{E}\left[\Sigma_{t+1} \,|\, \mathcal{F}_t\right] = \varrho\Sigma_t$. In particular, $\mathbf{E}[\Sigma_t] = \mathbf{E}[\Sigma_0]\varrho^t$ for all $t \in \mathbb{N}$. But $\Sigma_0 = \frac{n^2(\gamma-1)d_i^-}{m^2}$, where i denotes the mark of the root. As the latter is uniformly distributed on V, we deduce that $\mathbf{E}[\Sigma_0] = \frac{n(\gamma-1)}{m}$, and inserting into (25) completes the computation of $\mathbf{E}[(M_\star - M_t)^2]$. $\qquad\square$

Now that $M_\star$ is constructed, we may establish the representation announced in the introduction.

Lemma 16. *The random variable $M_\star$ admits the representation (6)-(7).*

Proof. Let $\mathcal{T}_\star^t(x)$ denote the set of all nodes y from which there is a child-to-parent path of length t to the node x in $\mathcal{T}_\star$. Writing $x_0, \dots, x_t$ for the corresponding path, we define

$$\mathbf{w}(y, x) \; := \; d_{i(x_0)}^- \prod_{k=0}^t \frac{1}{d_{i(x_k)}^+}.$$

Now, for every $t \in \mathbb{N}$ and every node x in the tree $\mathcal{T}_\star$, we define the quantity

$$Z_t(x) \; := \sum_{y \in \mathcal{T}_\star^t(x)} \mathbf{w}(y, x).$$

A martingale argument similar to the one above shows that the limit $Z_\star(x) = \lim_{t \to \infty} Z_t(x)$ exists almost-surely. By construction, we have for all $t \geq 0$,

$$M_{t+1} \; = \; \frac{n}{m} \sum_{x \to o} Z_t(x) \quad \text{and} \quad Z_{t+1}(x) \; = \; \frac{1}{d_{i(x)}^+} \sum_{y \to x} Z_t(y).$$

Passing to the limit, we deduce that almost-surely

$$M_\star \; = \; \frac{n}{m} \sum_{x \to o} Z_\star(x) \quad \text{and} \quad Z_\star(x) \; = \; \frac{1}{d_{i(x)}^+} \sum_{y \to x} Z_\star(y).$$

The first equation is precisely (6), since $i(o)$ is uniformly distributed on V and conditionally on $\mathcal{F}_0$, the random vector $(Z_\star(x), x \to o)$ consists of $d_{i(o)}^-$ i.i.d. coordinates. The second equation implies (7) since conditionally on $\mathcal{F}_0$, each $x \sim o$ satisfies the following: $i(x)$ has the out-degree distribution and conditionnally on $\mathcal{F}_1$, the random vector $(Z_\star(y), y \to x)$ consists of $d_{i(x)}^-$ i.i.d. coordinates whose marginal distribution is the same as $Z_\star(x)$. $\qquad \square$

8 Proof of Theorem 3

Thanks to (9), the proof of Theorem 3 boils down to establishing that for any fixed $t \in \mathbb{N}$,

$$2\|\pi_h - \pi_t\|_{\mathrm{TV}} \; \leq \; \sqrt{\frac{n\,(\gamma - 1)\,\varrho^t}{m(1 - \varrho)}} + o_{\mathbf{P}}(1).$$

We first prove that the left-hand side is concentrated around its expectation.

Lemma 17. *For any fixed $t \in \mathbb{N}$,*

$$\|\pi_h - \pi_t\|_{\mathrm{TV}} \; = \; \mathbf{E}\left[\|\pi_h - \pi_t\|_{\mathrm{TV}}\right] + o_{\mathbf{P}}(1).$$

Proof. Fix the environment. The probability that the random walk of length t starting from the in-degree distribution traverse a particular tail is trivially upper-bounded by the π_0-measure of the backward ball of radius t around its vertex, which is at most $\frac{\Delta^{t+2}}{m}$. Consequently, swapping two coordinates of the environment ω cannot alter the distribution π_t by more than $\frac{2\Delta^{t+2}}{m}$ in total variation. By the triangle inequality, we conclude that it cannot alter the quantity $Z = \|\pi_t - \pi_h\|_{\mathrm{TV}}$ by more than $b = \frac{2\Delta^{t+2}+2\Delta^{h+2}}{m}$. By a now classical application of the Azuma–Hoeffding inequality (see, e.g., [34, Section 3.2.]), this property implies that under the uniform measure, the random variable $\omega \mapsto Z(\omega)$ satisfies the following concentration inequality:

$$\mathbf{P}\left(|Z - \mathbf{E}[Z]| \geq \varepsilon\right) \quad \leq \quad 2\exp\left(-\frac{\varepsilon^2}{2mb^2}\right). \tag{26}$$

We may now let $n \to \infty$. With t fixed and h as in (8), we have $mb^2 \to 0$, as desired. $\qquad\square$

It only remains to bound the expectation. First, we may rewrite it as

$$2\mathbf{E}\left[\|\pi_t - \pi_h\|_{\mathrm{TV}}\right] \quad = \quad \mathbf{E}\left[|n\pi_t(\mathcal{I}) - n\pi_h(\mathcal{I})|\right],$$

where $\mathcal{I}$ is a uniformly chosen vertex, independent of the environment. Now, by Proposition 14, the total variation distance between the law of the random vector $(n\pi_t(\mathcal{I}))_{0\leq t\leq h}$ and that of $(M_t)_{0\leq t\leq h}$ is less than $p = \frac{\Delta^{2h+3}}{m}$. Since all coordinates are crudely bounded by Δ^{h+1}, we deduce that

$$\left|\mathbf{E}\left[|n\pi_t(\mathcal{I}) - n\pi_h(\mathcal{I})|\right] - \mathbf{E}\left[|M_t - M_h|\right]\right| \quad \leq \quad p\Delta^{h+1} \quad = \quad o(1).$$

Finally, the Cauchy–Schwarz inequality and the orthogonality of martingale increments yield

$$
\begin{aligned}
\mathbf{E}\left[|M_t - M_h|\right]^2 \quad &\leq \quad \mathbf{E}\left[(M_t - M_h)^2\right]\\
&\leq \quad \mathbf{E}\left[(M_t - M_\star)^2\right],
\end{aligned}
$$

and Proposition 15 concludes the proof.

9   Proof of Theorem 4

We first establish the following (weaker) result.

Lemma 18. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be non-expansive, i.e. $|f(x) - f(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$. Then,*

$$\left|\frac{1}{n}\sum_{i=1}^{n} f\left(n\pi_\star(i)\right) - \mathbf{E}[f(M_\star)]\right| \quad \xrightarrow[n\to\infty]{\mathbf{P}} \quad 0.$$

Proof. The assumption on f ensures that for any probability measures π, π' on V,

$$\left| \frac{1}{n} \sum_{i=1}^{n} f\left(n\pi(i)\right) - \frac{1}{n} \sum_{i=1}^{n} f\left(n\pi'(i)\right) \right| \leq 2\|\pi - \pi'\|_{\mathrm{TV}}. \tag{27}$$

Choosing $\pi = \pi_\star$, $\pi' = \pi_h$ and invoking (9), we see that

$$\frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_\star(i)\right) = \frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_h(i)\right) + o_{\mathbf{P}}(1).$$

Now, recall from the proof of Lemma 17 that swapping two coordinates of the environment cannot alter π_h by more than $\frac{2\Delta^{h+2}}{m}$ in total variation. In view of (27), we deduce that a swap cannot alter the variable $Z = \frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_h(i)\right)$ by more than $b = \frac{4\Delta^{h+2}}{m}$. Since $mb^2 \to 0$, the concentration inequality (26) implies

$$\frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_h(i)\right) = \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_h(i)\right)\right] + o_{\mathbf{P}}(1).$$

Observe that the expectation on the right-hand side is $\mathbf{E}\left[f\left(n\pi_h(\mathcal{I})\right)\right]$, where $\mathcal{I}$ is a uniform vertex, independent of the environment. Proposition 14 provides us with a coupling under which the random variables $n\pi_h(\mathcal{I})$ and M_h differ with probability less than $p = \frac{\Delta^{2h+3}}{m}$. Since both are bounded by Δ^{h+1} and since f is non-expansive, we obtain

$$\left| \mathbf{E}\left[\frac{1}{n} \sum_{i=1}^{n} f\left(n\pi_h(i)\right)\right] - \mathbf{E}\left[f\left(M_h\right)\right] \right| \leq p\Delta^{h+1} = o(1).$$

Finally, by the non-expansiveness of f, the Cauchy–Schwarz inequality, and Proposition 15,

$$\begin{aligned} \left|\mathbf{E}\left[f\left(M_h\right)\right] - \mathbf{E}\left[f\left(M_\star\right)\right]\right| &\leq \mathbf{E}\left[|M_h - M_\star|\right] \\ &\leq \sqrt{\mathbf{E}\left[(M_h - M_\star)^2\right]} \\ &\leq \sqrt{\frac{n(\gamma - 1)\varrho^h}{m(1 - \varrho)}} = o(1). \end{aligned}$$

Combining those four inequalities concludes the proof. $\qquad\square$

To conclude the proof of Theorem 4, we observe that $M_\star$ is bounded in L^2 uniformly in n by Proposition 15, and hence uniformly integrable as n varies. We may then apply the following general result. Write $\mathcal{P}_1(\mathbb{R})$ for the (Polish) space of Borel probability measures on $\mathbb{R}$ with finite first absolute moment, equipped with the Wasserstein distance $\mathcal{W}$.

Lemma 19 (Convergence of random measures in $\mathcal{P}_1(\mathbb{R})$)**.** *Let $(\mathcal{L}_n)_{n\geq 1}$ be a sequence of random elements of $\mathcal{P}_1(\mathbb{R})$, and let $(\mathbb{L}_n)_{n\geq 1}$ be a uniformly integrable sequence of deterministic elements of $\mathcal{P}_1(\mathbb{R})$. Assume that for every non-expansive function $f\colon \mathbb{R} \to \mathbb{R}$,*

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mathcal{L}_n - \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{L}_n \right| \xrightarrow[n\to\infty]{\mathbf{P}} 0. \tag{28}$$

Then $\mathcal{W}\left(\mathcal{L}_n, \mathbb{L}_n\right) \xrightarrow[n\to\infty]{\mathbf{P}} 0$.

Proof. It is classical that convergence in $\mathcal{P}_1(\mathbb{R})$ can be tested on a certain countable family $\mathcal{F}$ of non-expansive functions, in the following sense:

$$\mathcal{W}(\mathbb{L}_n, \mathbb{L}) \to 0 \iff \left(\forall f \in \mathcal{F}, \ \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{L}_n \to \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{L} \right). \tag{29}$$

One may take, for example, the function $x \mapsto |x|$ together with the functions $x \mapsto \frac{1-e^{-\theta [x]_+}}{\theta}$ and $x \mapsto \frac{1-e^{-\theta [x]_-}}{\theta}$ for positive rational numbers θ, since convergence in $\mathcal{P}_1(\mathbb{R})$ is equivalent to weak convergence together with convergence of the first absolute moment. We also recall the following standard result: a sequence of real-valued random variables $(X_n)_{n\geq 1}$ converges in probability to 0 if and only if from every subsequence $(X_{a(n)})_{n\geq 1}$, one can further extract a subsubsequence $(X_{a(b(n))})_{n\geq 1}$ converging to 0 almost-surely. Let us apply this to $X_n = \mathcal{W}(\mathcal{L}_n, \mathbb{L}_n)$: given an extractor a (i.e., an increasing function from $\mathbb{N}$ to $\mathbb{N}$), we will construct an extractor b such that

$$\mathcal{W}(\mathcal{L}_{a(b(n))}, \mathbb{L}_{a(b(n))}) \xrightarrow[n\to\infty]{\text{a.-s.}} 0. \tag{30}$$

First recall that uniform integrability means relative compactness in $\mathcal{P}_1(\mathbb{R})$: thus, the assumption ensures that there exists $\mathbb{L} \in \mathcal{P}_1(\mathbb{R})$ and an extractor c such that

$$\mathcal{W}(\mathbb{L}_{a(c(n))}, \mathbb{L}) \xrightarrow[n\to\infty]{} 0 \tag{31}$$

Combining this with assumption (28), we see that for each non-expansive function $f\colon \mathbb{R} \to \mathbb{R}$,

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mathcal{L}_{a(c(n))} - \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{L} \right| \xrightarrow[n\to\infty]{\mathbf{P}} 0.$$

In particular, for a fixed f, there is an extractor d such that

$$\left| \int_{\mathbb{R}} f \, \mathrm{d}\mathcal{L}_{a(c(d(n)))} - \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{L} \right| \xrightarrow[n\to\infty]{\text{a.-s.}} 0. \tag{32}$$

The extractor d depends on f, but by diagonal extraction one may construct one that satisfies (32) simultaneously for all f in the countable family $\mathcal{F}$. In view of (29), we conclude that

$$\mathcal{W}\left(\mathcal{L}_{a(c(d(n)))}, \mathbb{L}\right) \xrightarrow[n\to\infty]{\text{a.-s.}} 0.$$

Recalling the convergence (31), we see that the extractor $b(n) := c(d(n))$ satisfies (30). $\qquad\square$

References

[1] L. Addario-Berry, B. Balle, and G. Perarnau. Diameter and stationary distribution of random *r*-out digraphs. *ArXiv e-prints*, 2015.

[2] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

[3] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, pages 333–348, 1986.

[4] J. Barral. Moments, continuité, et analyse multifractale des martingales de Mandelbrot. *Probab. Theory Related Fields*, 113(4):535–569, 1999.

[5] J. Barral. Mandelbrot cascades and related topics. In *Geometry and analysis of fractals*, volume 88 of *Springer Proc. Math. Stat.*, pages 1–45. Springer, Heidelberg, 2014.

[6] A. Ben-Hamou and J. Salez. Cutoff for non-backtracking random walks on sparse random graphs. *ArXiv e-prints*, 2015.

[7] I. Benjamini, G. Kozma, and N. Wormald. The mixing time of the giant component of a random graph. *Random Structures Algorithms*, 45(3):383–407, 2014.

[8] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *ArXiv e-prints*, 2015.

[9] S. Chatterjee. Stein's method for concentration inequalities. *Probability theory and related fields*, 138(1):305–321, 2007.

[10] G.-Y. Chen and L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability*, 13(3):26–78, 2008.

[11] N. Chen, N. Litvak, and M. Olvera-Cravioto. Ranking algorithms on directed configuration networks. *ArXiv e-prints*, Sept. 2014.

[12] C. Cooper. Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Structural Information and Communication Complexity*, pages 1–14. 2011.

[13] C. Cooper and A. Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.*, 13(3):319–337, 2004.

[14] C. Cooper and A. Frieze. The cover time of random regular graphs. *SIAM J. Discrete Math.*, 18(4):728–740, 2005.

[15] C. Cooper and A. Frieze. The cover time of sparse random graphs. *Random Structures Algorithms*, 30(1-2):1–16, 2007.

[16] C. Cooper and A. Frieze. The cover time of the preferential attachment graph. *J. Combin. Theory Ser. B*, 97(2):269–290, 2007.

[17] C. Cooper and A. Frieze. The cover time of the giant component of a random graph. *Random Structures Algorithms*, 32(4):401–439, 2008.

[18] C. Cooper and A. Frieze. Stationary distribution and cover time of random walks on random digraphs. *J. Combin. Theory Ser. B*, 102(2):329–362, 2012.

[19] C. Cooper and A. Frieze. Vacant sets and vacant nets: Component structures induced by a random walk. *arXiv preprint arXiv:1404.4403*, 2014.

[20] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996.

[21] P. Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms*, 1(1):51–72, 1990.

[22] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981.

[23] J. Ding, E. Lubetzky, and Y. Peres. Mixing time of near-critical random graphs. *Ann. Probab.*, 40(3):979–1008, 2012.

[24] N. Fountoulakis and B. A. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms*, 33(1):68–86, 2008.

[25] D. A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, pages 100–118, 1975.

[26] H. Lacoin. The Cutoff profile for the Simple-Exclusion process on the circle. *ArXiv e-prints*, Feb. 2015.

[27] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times.* American Mathematical Soc., 2009.

[28] Q. Liu. The growth of an entire characteristic function and the tail probabilities of the limit of a tree martingale. In *Trees (Versailles, 1995)*, volume 40 of *Progr. Probab.*, pages 51–80. Birkhäuser, Basel, 1996.

[29] Q. Liu. Sur une équation fonctionnelle et ses applications: une extension du théorème de Kesten-Stigum concernant des processus de branchement. *Adv. in Appl. Probab.*, 29(2):353–373, 1997.

[30] Q. Liu. On generalized multiplicative cascades. *Stochastic Process. Appl.*, 86(2):263–286, 2000.

[31] Q. Liu. Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic Process. Appl.*, 95(1):83–107, 2001.

[32] E. Lubetzky and Y. Peres. Cutoff on all Ramanujan graphs. *ArXiv e-prints*, 2015.

[33] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

[34] C. McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.

[35] A. Nachmias and Y. Peres. Critical random graphs: diameter and mixing time. *Ann. Probab.*, 36(4):1267–1286, 2008.

[36] Y. Peres. American institute of mathematics (AIM) research workshop "sharp thresholds for mixing times" (Palo Alto, December 2004). *Summary available at http://www.aimath.org/WWN/mixingtimes*.

[37] U. Rösler. A fixed point theorem for distributions. *Stochastic Process. Appl.*, 42(2):195–214, 1992.

[38] L. Saloff-Coste. Random walks on finite groups. In *Probability on discrete structures*, pages 263–346. Springer, 2004.

[39] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.